

# Interactive reinforcement learning innovation to reduce carbon emissions in railway infrastructure maintenance

Sresakoolchai, Jessada; Kaewunruen, Sakdirat

DOI:

[10.1016/j.dibe.2023.100193](https://doi.org/10.1016/j.dibe.2023.100193)

License:

Creative Commons: Attribution (CC BY)

*Document Version*

Peer reviewed version

*Citation for published version (Harvard):*

Sresakoolchai, J & Kaewunruen, S 2023, 'Interactive reinforcement learning innovation to reduce carbon emissions in railway infrastructure maintenance', *Developments in the Built Environment*, vol. 15, 100193. <https://doi.org/10.1016/j.dibe.2023.100193>

[Link to publication on Research at Birmingham portal](#)

## General rights

Unless a licence is specified above, all rights (including copyright and moral rights) in this document are retained by the authors and/or the copyright holders. The express permission of the copyright holder must be obtained for any use of this material other than for purposes permitted by law.

- Users may freely distribute the URL that is used to identify this publication.
- Users may download and/or print one copy of the publication from the University of Birmingham research portal for the purpose of private study or non-commercial research.
- User may use extracts from the document in line with the concept of 'fair dealing' under the Copyright, Designs and Patents Act 1988 (?)
- Users may not further distribute the material nor use it for the purposes of commercial gain.

Where a licence is displayed above, please note the terms and conditions of the licence govern your use of this document.

When citing, please reference the published version.

## Take down policy

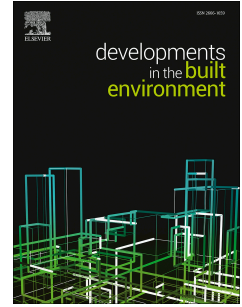
While the University of Birmingham exercises care and attention in making items available there are rare occasions when an item has been uploaded in error or has been deemed to be commercially or otherwise sensitive.

If you believe that this is the case for this document, please contact [UBIRA@lists.bham.ac.uk](mailto:UBIRA@lists.bham.ac.uk) providing details and we will remove access to the work immediately and investigate.

# Journal Pre-proof

Interactive reinforcement learning application to reduce carbon emissions in railway infrastructure maintenance

Jessada Sresakoolchai, Sakdirat Kaewunruen



PII: S2666-1659(23)00075-3

DOI: <https://doi.org/10.1016/j.dibe.2023.100193>

Reference: DIBE 100193

To appear in: *Developments in the Built Environment*

Received Date: 15 May 2023

Revised Date: 30 June 2023

Accepted Date: 1 July 2023

Please cite this article as: Sresakoolchai, J., Kaewunruen, S., Interactive reinforcement learning application to reduce carbon emissions in railway infrastructure maintenance, *Developments in the Built Environment* (2023), doi: <https://doi.org/10.1016/j.dibe.2023.100193>.

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2023 Published by Elsevier Ltd.

# 1 Interactive reinforcement learning application to reduce carbon 2 emissions in railway infrastructure maintenance

3 Jessada Sresakoolchai and Sakdirat Kaewunruen\*

4 *School of Civil Engineering, University of Birmingham, B15 2TT, Birmingham, United Kingdom*

5 \*Email: [s.kaewunruen@bham.ac.uk](mailto:s.kaewunruen@bham.ac.uk)

6 \*Tel.: +44 (0) 121 414 2670

7

## 8 HIGHLIGHTS

- 9 • The first reinforcement learning model for railway carbon emissions reduction
- 10 • Field data robustly enables the creation of customized environments for the model
- 11 • Environment's states are obtained from defective track geometry and track  
12 component
- 13 • A complex combination of maintenance activities is adopted as an action space
- 14 • Reduce the defect and carbon emissions using an interactive and dynamic approach

15

16 ABSTRACT: Carbon emission is one of the primary contributors to global warming. The global  
17 community is paying great attention to this negative impact. The goal of this study is to reduce  
18 the negative impact of railway maintenance by applying reinforcement learning (RL) by  
19 optimizing maintenance activities. Railway maintenance is a complex process that may not be  
20 efficient in terms of environmental aspect. This study aims to use the potential of RL to reduce  
21 carbon emission from railway maintenance. The data used to create the RL model are  
22 gathered from the field data between 2016-019. The study section is 30 kilometers long.  
23 Proximal Policy Optimization (PPO) is applied in the study to develop the RL model. The results  
24 demonstrate that using RL reduces carbon emission from railway maintenance by 48%, which  
25 generates a considerable amount of carbon emission reduction and reduces railway defects  
26 by 68%, which also improves maintenance efficiency significantly.

27 Keywords: reinforcement learning, carbon emission, railway system, maintenance, railway  
28 defects, environmental impact

## 29 1 INTRODUCTION

30 Carbon emission is currently one of the major drivers of catastrophic global warming and  
31 climate change. Climate change is caused by the accumulation of greenhouse gases such as  
32 carbon dioxide in the atmosphere [1]. These gases trap heat from the sun and cause the  
33 Earth's surface temperature to rise. This leads to a range of negative impacts such as more  
34 frequent and severe events [2, 3]. As a result, carbon emission is garnering increased  
35 attention. Worldwide communities then attempt to diminish it in order to minimize  
36 environmental impacts. Rail transportation is one of the most environmentally friendly ways  
37 of transportation [4]. However, activities in the railway system contribute to carbon emission.  
38 Because railway projects have a long service life, the operating and maintenance phases also  
39 contribute to carbon emissions about 6% [5] of the total carbon emission. As a result, reducing  
40 carbon emissions in railway activities will have a substantial impact on the environment.

41 Railway maintenance may be done in several methods, including corrective maintenance,  
42 preventive maintenance, and predictive maintenance (condition-based maintenance) [6]. In  
43 brief, corrective maintenance is performed when something fails. Preventive maintenance  
44 tends to be routine maintenance when the maintenance is performed although there is  
45 nothing fails. Predictive maintenance is an approach for planning maintenance based on the  
46 current condition of components. Predictive maintenance appears to be the most reasonable  
47 alternative for doing maintenance operations nowadays since it performs just what is  
48 required to preserve railway infrastructure in acceptable conditions. Maintenance is

49 scheduled according to the existing state of each railway component and section.  
50 Maintenance will be undertaken only if there is a risk of failure or some values reach  
51 thresholds. However, the main challenge of applying predictive maintenance is it needs a  
52 reliable tool for predicting and planning. It can be seen that optimal decision-making will  
53 result in minimum defect and cost [7]. Fortunately, there are many machine learning  
54 techniques that are being developed and computational power become more powerful  
55 compared to the past decade, the application of predictive maintenance becomes  
56 increasingly feasible.

57 Nowadays, deep reinforcement learning (RL) is employed to tackle a wide range of problems.  
58 However, its use in railway maintenance is currently restricted. Through model training, an  
59 agent in the RL model which is used as a representative will learn to maximize rewards or limit  
60 penalties. In this study, carbon dioxide (CO<sub>2</sub>) is selected as a representative of greenhouse  
61 gas because CO<sub>2</sub> is the most abundant greenhouse gas (65%) used to determine rewards for  
62 the RL agent along with defects. During the training process, maintenance activities will be  
63 performed by the agent to minimize the defect occurrence. However, maintenance activities  
64 also create CO<sub>2</sub> which is calculated based on the energy and material used for each  
65 maintenance activity. Therefore, the agent has to perform maintenance activities to minimize  
66 occurring defects while the agent has to limit the number of maintenance activities as  
67 necessary to limit the amount of CO<sub>2</sub> as well. The amount of carbon emission is used as a  
68 reward.

69 The aim of this study is to use RL to minimize carbon emission from railway maintenance  
70 activities. The scope of this study will be limited to railway maintenance activities. This study

71 will apply PPO to create a deep RL model because it produces many superior outcomes while  
72 it is more stable. PPO requires the least amount of time for training [8]. Field data are applied  
73 to create the RL model using Monte Carlo simulation and further analysis. The customized  
74 environment of the reinforcement learning model is created based on the real characteristic  
75 of the railway system. Normal distribution and Monte Carlo simulation are used to code the  
76 environment to generate the following step when the agent takes action. Track geometry  
77 parameters and component defect probabilities will be improved or deteriorated based on  
78 the actions the agent takes. For example, when the agent chooses to perform tamping, some  
79 parameters will be improved. This information is based on the field data. However, it is worth  
80 noting that there are multiple options for the agent to take action because there are seven  
81 maintenance activities that the agent can take and each maintenance activity is independent.  
82 A further description is presented in 3.2. The customized environment of RL is created to meet  
83 the challenges of this study. The study's contributions include that the created RL model can  
84 be employed by railway operators to more efficiently schedule railway maintenance  
85 operations when carbon emission from maintenance activities is decreased, as is the number  
86 of railway defects. As a result of fewer defects in railway networks, it will result in a more  
87 environmentally friendly railway system, and improved serviceability, trustworthiness, and  
88 safety [9, 10]. Furthermore, railway operators can apply the created RL model to support  
89 decision-making for railway maintenance. This conforms to the challenge statement of  
90 Network Rail [10] which aims to improve the technical strategy in different aspects such as  
91 asset management, social and environmental, maintenance, operation, and cost  
92 management. For asset management, important aspects are Reliability, Availability,  
93 Maintainability, Safety (RAMS). RAMS is an important tool to improve the availability of the

94 system. RAMS has to be applied from the beginning of the project. At the same time, it should  
95 be updated along the project to ensure good asset management. This study also aims to  
96 improve RAMS in the railway system by developing the RL model to improve the reliability of  
97 the system by decreasing defects, availability by keeping tracks free from defects,  
98 maintainability by decreasing the severity of defects, and safety will be a result of the good  
99 quality track. It can be seen that the RL model that will be developed in this study will have  
100 benefits for asset management not limited to the environmental aspect only. This study's  
101 novelty is it is the first paper using RL to reduce carbon emission in railway maintenance,  
102 which has never been done before. Furthermore, the proposed RL approach is unique since  
103 it is based on a customized environment that is tailored to the problem.

## 104 2 LITERATURE REVIEW

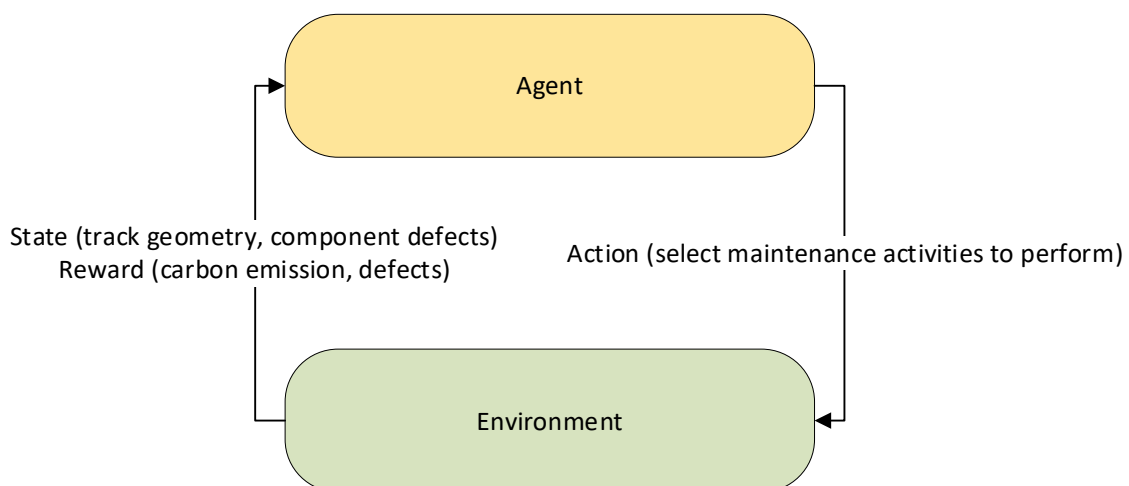
105 RL was initially presented in the early 1980s. It was created to address issues such as how to  
106 respond or what to do in certain scenarios [11]. The goals of varied responses are to maximize  
107 rewards or consequences from actions. An agent is utilized and trained as a representative to  
108 know how to respond to various scenarios. The agent is not instructed what to do, rather it  
109 must discover for itself how to maximize rewards at the final stage. A problem is that each  
110 action influences not just the immediate rewards but also what happens following the  
111 timestep (stage) and the total rewards at the final stage. All of these are crucial features of RL  
112 that other types of machine learning approaches lack and cannot address. For more detail, RL  
113 can be used to solve problems continuously and it uses information from previous stages to  
114 take actions from the following stages. Other categories of machine learning do not have this  
115 capability because they perform prediction only one time in an epoch. For example,  
116 supervised learning will receive a set of features and predict labels. Unsupervised learning will

117 discover the insights from the data without labeling the dataset. However, RL receives a set  
118 of features or states and the agent will take action based on the states. The environment of  
119 the RL model will also generate a new stage based on the action and these processes will  
120 repeat until the end of the training. An example of RL application in the railway industry is  
121 train rescheduling to minimize delay by regulating the movements of rolling stock in real-time  
122 and continuously [12].

123 Chess and other game players, industrial robots, and pilot assistance in passenger  
124 automobiles are all examples of prominent RL applications [11]. To tackle RL difficulties, a  
125 variety of strategies are employed. The simplest of the initial approaches is the Tabular  
126 Solution Method. The idea for this approach came from the trial-and-error process. Every  
127 feasible stage from various actions is represented in tabular form. When all possibilities are  
128 provided, the RL agent will know what the optimum action is when dealing with different  
129 stages. However, it can be observed that the crucial drawback is that this approach can only  
130 be used when the options of stages and actions are restricted. In other words, the number of  
131 stages and actions are clear and known such as Tic-Tac-Toe where every possible stage is  
132 known. However, this approach is not suitable for the problem in this study because the  
133 railway system is complicated and the possibility of states is unlimited. Otherwise, it would  
134 be impossible to describe all feasible stages and appropriate actions. Other approaches, such  
135 as Markov Decision Processes, Dynamic Programming, or Monte Carlo Methods, were  
136 created to be utilized with RL to solve this constraint and have the potential to solve problems  
137 with unlimited possibilities such as chess or other games. Many strategies have been created  
138 over the last decade to extend the potential of RL, and other current RL models have been  
139 established and demonstrated to be more effective.



140 The first terminology to be noted is a stage. Stages are various environmental conditions at  
141 different timesteps. In each stage, two components, the agent and the environment,  
142 constantly interact. The environment will provide information to the agent via stages and  
143 rewards. The agent will then respond by taking action against the environment. Following  
144 that, the environment will give the agent new stages and rewards. The available actions are  
145 categorized into action spaces. This procedure will be repeated till the training is completed.  
146 Figure 1 depicts a flowchart of this procedure. A reward is given to the agent at each stage  
147 based on how successfully the agent reacts to the environment. At the end of the training,  
148 rewards such as win-loss results may be awarded. This might vary based on the problem being  
149 attempted to tackle. This information will be used to choose the optimal policy for the agent.  
150 It is worth noting that the policy in this study refers to a strategy that an agent uses to take  
151 action in the environment. It defines how the agent selects actions based on the states of the  
152 environment. A policy can be different depending on the RL algorithms and problems. The  
153 goal of RL is to find an optimal policy that maximizes the rewards or long-term performance  
154 in the environment.



155

156

Figure 1 Flowchart of the RL model

157 RL has become more prominent in recent years. It has been used in a variety of domains [13],  
158 including communication and networking [14], biology [15], electrical systems [16], robotics  
159 [17], transportation [18], medical [19], finance [20], or engineering [21]. Following are some  
160 instances of RL applications.

161 In the railway industry, RL has been used in a variety of ways. Sedghi et al. [22] conducted a  
162 literature review on this subject. The majority of the techniques in their assessment were  
163 based on mathematics or probabilistic approaches such as stochastic modal, mixed integer  
164 programming, simulation, Markovian model, and machine learning, which are popular today  
165 and deliver satisfactory performance. The stochastic modal considers the uncertainty of the  
166 environment. Instead of selecting a single deterministic action for each state, a stochastic  
167 policy assigns a probability distribution over the available actions based on the states. Mixed  
168 integer programming is a mathematical optimization technique that is used to optimize the  
169 policies of agents. The Markovian model is a fundamental mathematical framework used in  
170 RL to model and solve sequential decision-making problems. It captures the future state and  
171 reward depending on the current state and action based on the past stages. Machine learning  
172 is the most advanced approach because agents are trained multiple times which allows agents  
173 to learn from the environment to have abilities to choose the best action under certain  
174 situations.

175 Šemrov et al. [23] used Q-Learning to reschedule single-track trains. The agent's action spaces  
176 were comprised of two actions: stop and go. The delay was used to compute rewards. The  
177 model was tested with a three-station scenario. It was discovered that utilizing RL might help  
178 to lessen railway delays. Cui et al. [24], Khadilkar [25], and Zhu et al. [12] also corroborated

179 this observation. Other railway sector uses of RL include control [26], power management  
180 [27-29], inspection [30], and alignment optimization [31].

181 RL has not been widely used in railway maintenance. Only one research has been conducted  
182 by Mohammadi and He [32]. Deep Q-learning (DQN) was used to create a decision-making  
183 tool for railway maintenance. This method used the track quality index (TQI) and the hazard  
184 index as inputs. These two indications were obtained using RL. These two were then utilized  
185 to train the RL model. Their action spaces included five activities: preventative tamping,  
186 preventive grinding, condition-based tamping, condition-based grinding, and renewal. When  
187 compared to the baseline, they discovered that using the suggested technique might  
188 minimize the TQI and hazard index. However, the study's research gap is that they employed  
189 summary indices such as TQI and hazard indexes instead of precise track geometry  
190 parameters. Previous years' maintenance tasks were not included, and the action spaces  
191 might be expanded for greater comprehensiveness. Furthermore, new sophisticated or  
192 cutting-edge strategies should be tested in order to increase the overall effectiveness of the  
193 RL model.

194 According to the literature review, the use of RL for railway infrastructure maintenance and  
195 carbon emission reduction is still in its early stages, with relatively little research available.  
196 There are gaps in research in this area that can be filled. For example, each track geometry  
197 parameter may be taken into account for more practical, comprehensive maintenance  
198 operations that can be incorporated into the RL model, or field data can be linked with the RL  
199 model. As a result, the goal of this study is to address as many of these gaps as possible by  
200 creating a method for using RL to reduce carbon emission from railway infrastructure

201 maintenance based on each track geometry parameter and individual track component  
202 defect. To guarantee that the produced model mimics the real-world condition as closely as  
203 possible, detailed maintenance operations and filed data are part of the RL model  
204 development.

### 205 3 METHODOLOGY

#### 206 3.1 *RL model and Proximal Policy Optimization (PPO)*

207 RL is one of the three major types of machine learning besides supervised and unsupervised  
208 learning. RL is gaining popularity at the moment. Agents in RL models are trained and learn  
209 from their environment [33]. Environments feature rules that agents must comply with, such  
210 as limitations and available actions. Agents will interact with their environments based on  
211 their states ( $s_t \in S$  where  $S$  is possible states), which can be a discrete timestep ( $t =$   
212  $1, 2, 3, \dots, n$ ), by completing actions ( $a_t \in a(s_t)$ ) in states ( $s_t$ ) where  $a(s_t)$  feasible action  
213 exists. Following that, agents will be rewarded (or penalized  $R_t$ ) for their actions [32]. The  
214 states will then transition to the next state, and the agents will conduct their acts once more.  
215 This method will be repeated until the training or environment states are completed. The goal  
216 of agent training is to increase rewards or decrease penalties.

217 PPO will be used in this study to create the RL model for a variety of reasons. PPO is by far the  
218 most highly sophisticated and best algorithm. It is designed to be more stable than other  
219 policy gradient methods, making it easier to train agents in complex environments. It achieves  
220 the greatest reward in the shortest number of steps and with the least instability. It can learn  
221 effectively with fewer samples and less fine-tuning than other policy gradient methods. In  
222 terms of flexibility, PPO can be applied to a wide range of tasks and environments, including

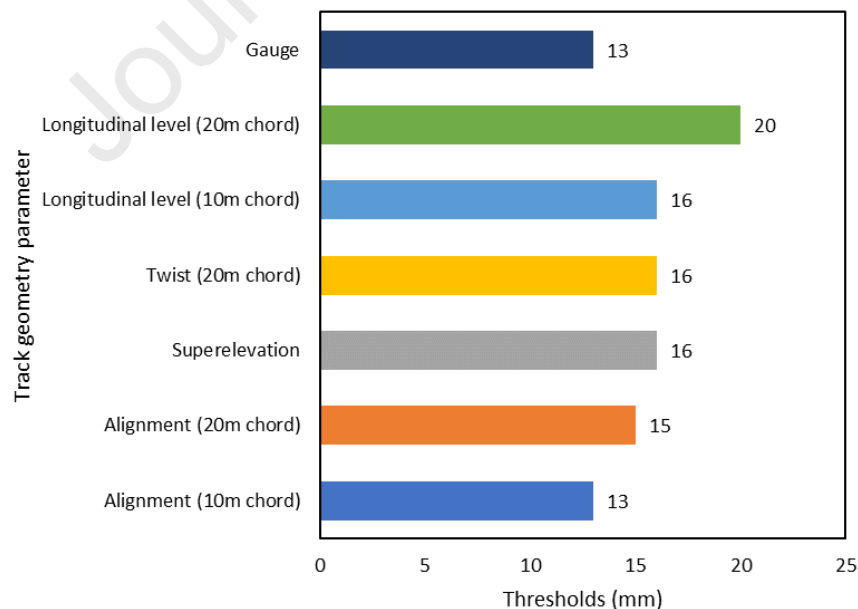
223 those with continuous or high-dimensional action spaces. Furthermore, in comparison to  
224 other RL methods, the training period is quite short [8, 34]. PPO is designed to find an optimal  
225 policy by updating the policy using data collected from interactions with an environment. In  
226 the beginning, PPO will collect data by executing the current policy in the environment. The  
227 information that is collected in this step is states, actions, rewards, and other relevant  
228 information. Then, it will compute how much better or worse an action is compared to the  
229 average action at a particular state. This is used to update policies. Then, PPO will update the  
230 policy by performing multiple epochs of policy updates using the collected data. In each step,  
231 the policy is updated to maximize the objective function. These processes are repeated again  
232 and again until the convergence is achieved. As a result of its performance, advancement, and  
233 novelty, PPO has been deemed the best methodology in this study.

### 234 3.2 *Data Characteristics and Preparation*

235 MRS Logística S.A. has provided field data for a 30-km railway section from 2016 to 2019. The  
236 data comes from a variety of sources, including track geometry measurements, rail  
237 component defect inspection reports, and actual maintenance records for the heavy haul rail  
238 networks.

239 Track geometry cars gather foot-by-foot track geometry parameters to obtain track geometry  
240 parameters or the sampling rate is 100 Hz. Superelevation, longitudinal level (10m chord),  
241 longitudinal level (20m chord), alignment (10m chord), alignment (20m chord), gauge, and  
242 twist (20m chord) are all included in the data sets. Seven track geometry parameters will be  
243 employed as inputs to the RL model. In other words, they will be utilized in the states specified  
244 by the agent to determine the following actions. The reason for using the track geometry

245 parameter is this is the fundamental information that every railway operator has. Moreover,  
246 this information directly represents the quality of the track. Thresholds are established by the  
247 company as four priorities. Priority 1 indicates that the track geometry parameters are  
248 extremely poor and that track sections should be maintained as soon as possible, whilst  
249 priority 4 indicates that track sections should be included in the normal maintenance  
250 schedule. In this scenario, priority 4, the least worried trigger level, will be selected as the  
251 threshold for the RL model to consider rewards and penalties since the goal of this study is to  
252 execute a predictive maintenance approach that keeps the track free of defects or limit them  
253 to a minimum. Figure 2 depicts the threshold for each track geometry parameter which is the  
254 base operating conditions (BOCs) according to MRS Logística S.A. Track sections with track  
255 geometry parameters that surpass the threshold are regarded as defective, and the number  
256 of defects is determined by the number of exceeding track geometry values.



257

258

Figure 2 Track geometry parameter thresholds [35]

259 The following data source is defect inspection reports, which gather various track component  
260 defects. There are 71 various forms of track component defects, which vary according to the  
261 track component. There are different types of component defects such as broken rails, broken  
262 frogs, or missing sleepers. To simplify the analysis, they are divided into five groups depending  
263 on track components: ballast, fastener, rail, sleeper, and switch and crossing. If defects are  
264 discovered, certain track sections are deemed defective, and the total number of defects is  
265 determined by the number of track component defects. When track geometry and  
266 component defects are combined, there are a total of 12 defect categories: seven for track  
267 geometry defects and five for track component defects.

268 The final information source is maintenance records. Tamping, rail grinding, ballast cleaning,  
269 sleeper replacement, rail replacement, fastening component replacement, and ballast  
270 unloading are the seven maintenance activities listed. The RL model's action space will be  
271 these seven maintenance activities. In this scenario, the action space may be thought of as  
272 seven binary spaces in which each maintenance activity can be performed or not performed.  
273 It is worth emphasizing that, in practice, maintenance activities are more sophisticated  
274 because they may be performed individually. They can be mixed in any number from 0 to 7.  
275 As a result, the probability combination principle is used to consider the alternative actions in  
276 each state of the RL model. Equation 1 is used to determine the total combination with  
277 repetition, where  $n$  is the number of alternatives and  $r$  is the size of the combination.  $n$  equals  
278 seven in the equation, and  $r$  can be modified from zero to seven. The total number of possible  
279 actions or combinations is 128. There are almost one million sets of data from maintenance  
280 records that could be regarded as thorough enough to analyze maintenance activities,

281 changes in track geometry parameters, and the incidence of track component defects in  
282 normal distribution patterns.

$$Total\ combinations = \frac{n!}{(n-r)!r!} \quad \text{Equation 1}$$

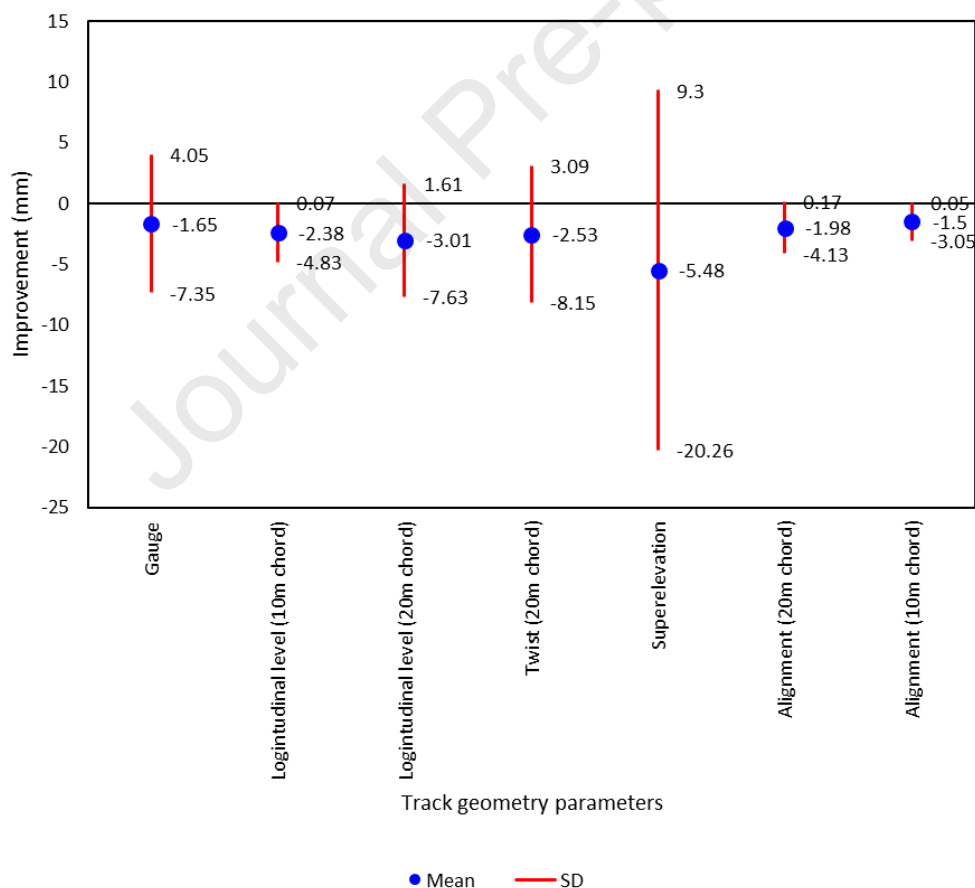
283 Every data source is integrated to combine and process to be prepared as the preliminary  
284 input of the RL model. These data will be utilized to create a customized RL environment.  
285 Different maintenance activities can be integrated with each stage, resulting in a high degree  
286 of variation in maintenance activities undertaken. Nevertheless, the datasets are abundant  
287 enough to enable this heterogeneous action space through numerical processing. In other  
288 words, the data is large enough to predict how much track geometry properties would  
289 improve or degrade when maintenance activities are carried out. In this case, the lower  
290 number is preferable. The size of the degradation and improvement are considered based on  
291 the field data. The relationships between the change in degradation and improvement are  
292 based on the maintenance activities performed. For example, when tamping is performed,  
293 the track geometry parameters will improve and the size of improvement is based on the field  
294 data from track geometry measurement and maintenance report. Simultaneously, how much  
295 the possibility of each track component defect will rise or diminish when alternative  
296 maintenance actions based on the same principle are done. Some examples for  
297 demonstrating a clearer view are given following.

298 As mentioned, seven maintenance activities. These activities have been collected from the  
299 maintenance report which covers the maintenance of the track structure. Figure 3 and Figure  
300 4 demonstrate examples of improvements and degradation in track geometry parameters



301 when tamping is performed or is not performed in the context of a normal distribution. It  
302 should be noted that the positive numbers represent the deterioration while the negative  
303 numbers represent the improvement because the low-track geometry parameters are  
304 desirable. As representations, the mean and standard deviation (SD) are utilized. The figures  
305 show only two simple examples that are being analyzed. There are additionally 127 more  
306 instances based on potential actions that will be utilized to create the relationships between  
307 actions and states in the RL model. Figure 5 also illustrates track component defects as an  
308 example of the likelihood's change when component defects are not identified and identified  
309 when rail grinding is conducted and not conducted respectively. From the figure, the chance  
310 of defects to occur tends to decrease when the maintenance is performed and increase when  
311 the maintenance is not performed. However, these numbers of probability are  
312 mathematically limited to be in the range of zero to one. This is also included in the code to  
313 create the customized environment for the RL model which will be described in the following  
314 section. The figure only represents a simple scenario of the track component defects. In  
315 reality, there are more intricate combinations of maintenance activities that impact the track  
316 component defect incidence that is analyzed and used as inputs in the RL model in forms of  
317 the association between actions and states. The most complicated issue is determining how  
318 to update the states when the agent conducts actions. In this scenario, the states are divided  
319 into twelve sub-states that represent track geometry parameters and track component defect  
320 occurrences. States are real numbers that indicate the extent of abnormalities in track  
321 geometry parameters while they are binary integers that indicate whether or not a track  
322 component defect exists. Changes in each state are evaluated while updating the states of  
323 the RL model utilizing field data including track geometry measurements, defect inspection

324 reports, and maintenance records as mentioned previously. Changes in states are based on  
 325 this data using a normal distribution associated with the aforementioned specified  
 326 maintenance actions. Now that all of the necessary data has been processed. The  
 327 environment for the RL model can be created. The reason for using normal distribution  
 328 without considering time-series data of the change in track geometry parameter is because  
 329 the data are not continuous and they depend on different performed maintenance activities.  
 330 Therefore, the transition model has a limitation to be applied with the reinforcement learning  
 331 model in this study.

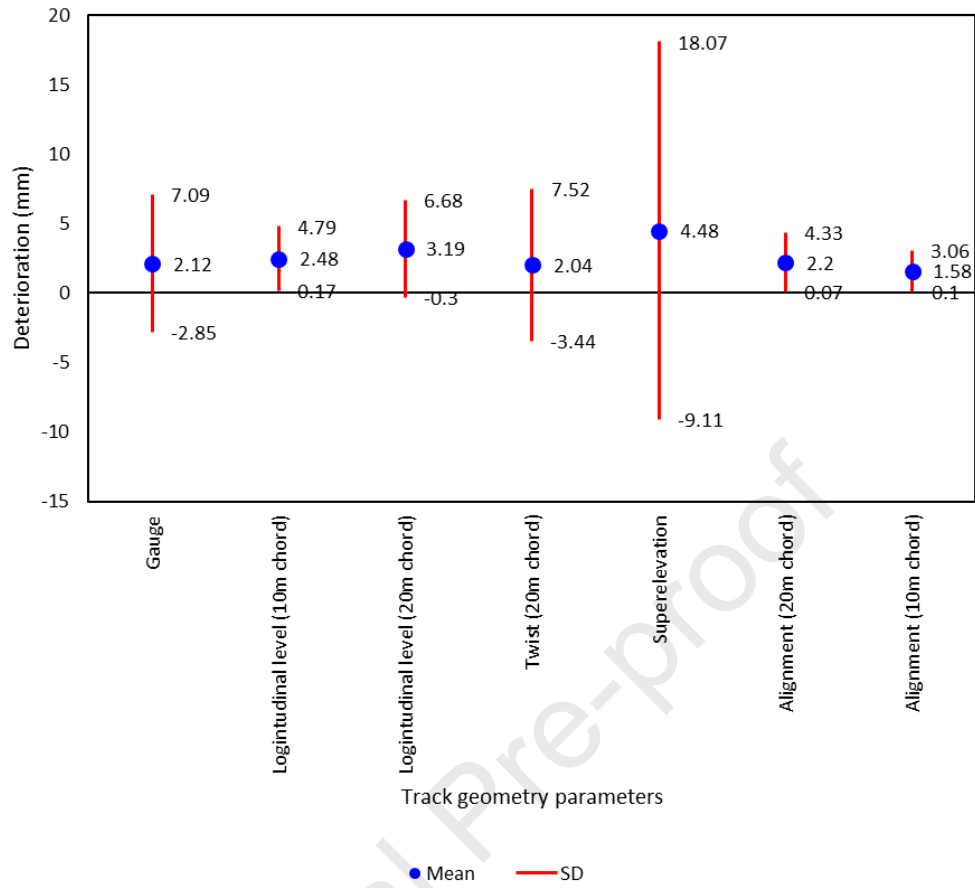


332

333

Figure 3 Track geometry parameters' improvement when tamping is performed

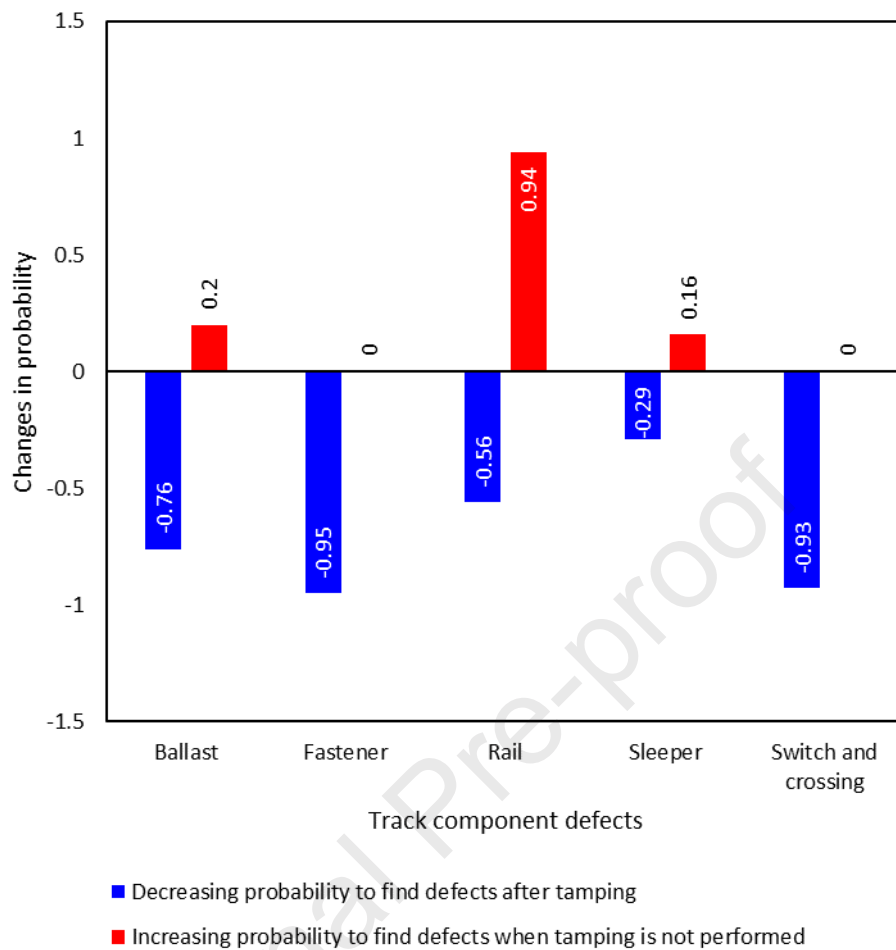
334



335

336

Figure 4 Track geometry parameters' deterioration when tamping is not performed



337

338

Figure 5 Relationship between defect occurrence and rail grinding

339 3.3 *Problem description and customized environment*

340 To achieve the purpose to train the RL agent, the problem in this study is unique. Therefore,

341 a customized environment has to be created to match the problem in this study. To create

342 the customized environment, OpenAI Gym is the employed platform. In addition, Stable

343 Baselines are used to develop the RL agent. The first step of the customized environment

344 creation is to define the action space and observation space. The dimension of the action

345 space and observation space have to correspond to the number of available actions and

346 states. In this case, the dimension of the action space will be seven (types of maintenance

347 activities) and the dimension of the observation space will be 12 (values to consider defects).

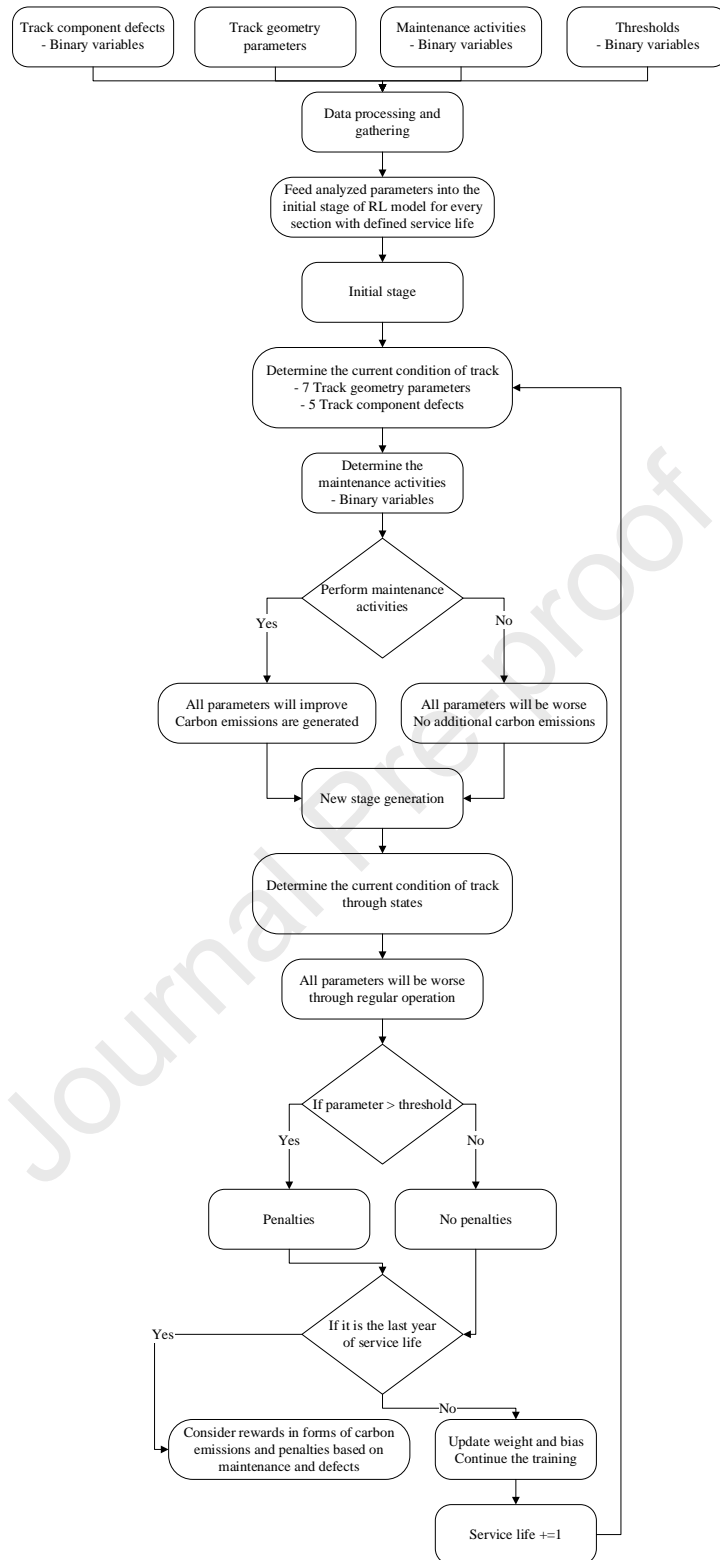
348 In the study, the action space is discrete when it is in the form of multi-binary (do or not do)

349 while the observation space is continuous or the box (real number representing track  
350 geometry parameters and probability of component defect occurrence). The step is used to  
351 represent each stage or timestep in the environment. The reset function is used to reset the  
352 environment to the initial step and retrain the RL agent because the agent will be trained  
353 multiple times. states are variables that are collected in the observation space which the  
354 agent uses as information to choose the next action. Rewards can be defined in each step or  
355 at the final step of the training. Rewards can be defined based on states. The done function  
356 is used to tell the agent that it reaches the final step of the training and the environment has  
357 to be reset to train the agent again.

358 The agent will learn how to perform maintenance activities based on 12 states which include  
359 seven track geometry parameters and five track component defects. There are five  
360 maintenance activities as available combined actions so 128 different maintenance actions  
361 are feasible. Varied combinations of maintenance actions result in different improved and  
362 degraded track geometry characteristics, as well as the chance of track component defect  
363 incidence, as determined by field data analysis. The RL agent tries to decrease the carbon  
364 emission generated from maintenance activities while maintaining tracks free from defects.  
365 A non-defective track means that the track section has no track component defects and no  
366 track geometry parameters that exceed established criteria. Because this study has not  
367 included the maintenance cost in the training process of the agent, the agent selects the  
368 maintenance activities to perform based on the carbon emissions and the current status of  
369 defects. To avoid the bias of the agent to select only maintenance activities that cause small  
370 carbon emissions, the rewards (penalties) based on defects are set to have high values to

371 ensure that the agent will try to prevent the occurring defects instead of minimizing carbon  
372 emissions.

373 Figure 6 demonstrates the complete process of the RL model. Every parameter in the first  
374 state is defined based on the field data. The agent must next take action by deciding which  
375 maintenance tasks are required to be performed. Following the action, the environment will  
376 respond by producing a set of new states that take into account the right values of each state  
377 depending on the previously given field data to ensure that it is adequate for real-world  
378 applications. This method will be repeated till the training is completed. Penalties from two  
379 categories are used to determine rewards: carbon emission and defect incidences. Track  
380 geometry defects are determined through Figure 2, whereas track component defects are  
381 determined by occurrence. Because agent training strives to limit carbon emission while  
382 maintaining the track defect-free, the penalties for defect incidence in this study are  
383 particularly substantial compared to the penalties based on carbon emission. Loss and policy  
384 entropy are utilized to demonstrate the performance of the RL model. The loss indicates the  
385 model's deviation from expectation, while the policy entropy shows how successfully the  
386 agent responds to challenges. The desired values for these two parameters are both 0. During  
387 the training process, the timestep is set to be 100 years as the service life of the project which  
388 should be comprehensive in many cases of railway projects that the service life can be ranged  
389 from 50 to 70 years. In addition, because the maintenance plan is prepared on an annual  
390 basis, therefore, one step of the RL model represents one year of the operation and  
391 maintenance stage of the railway project.



392

393

Figure 6 The RL model's workflow

394 For the determination of carbon emission, the study will limit carbon emission to

395 maintenance activities and employed materials solely. For the carbon emission created by

396 maintenance operations, two sources of carbon emission are considered: material  
 397 consumption and electricity utilized to operate the equipment or machine. Some procedures  
 398 may solely involve materials since a machine is not necessary. At the same time, some  
 399 maintenance procedures may not necessitate the use of materials, resulting in carbon  
 400 emission based only on the energy used to operate the machine. Table 1 shows the  
 401 calculation in further detail. As previously stated, the total carbon emission of various  
 402 maintenance activities will be supplied to the RL to train the agent. From the table, the speed  
 403 and power information are gathered from different sources such as specifications or catalogs.  
 404 Then, the energy consumption is calculated based on the speed and power of the machine  
 405 and the length of the track section. In this case, the length of a track section is one foot where  
 406 the carbon emission from electricity generation is 0.509 (kg CO<sub>2</sub>e) [36, 37]. Then, the carbon  
 407 emission based on the material is calculated. For ballast, the carbon emission from the  
 408 production is 3.4 kg CO<sub>2</sub>e/ton [38]. The sleeper production generates a carbon emission of  
 409 0.124 kg CO<sub>2</sub>e/kg [39] and assumes that the sleeper weight is 70 kg. Last, the carbon emission  
 410 from steel production is 1.85 kg CO<sub>2</sub>e/kg [40].

411 Table 1 Carbon emission from maintenance activities

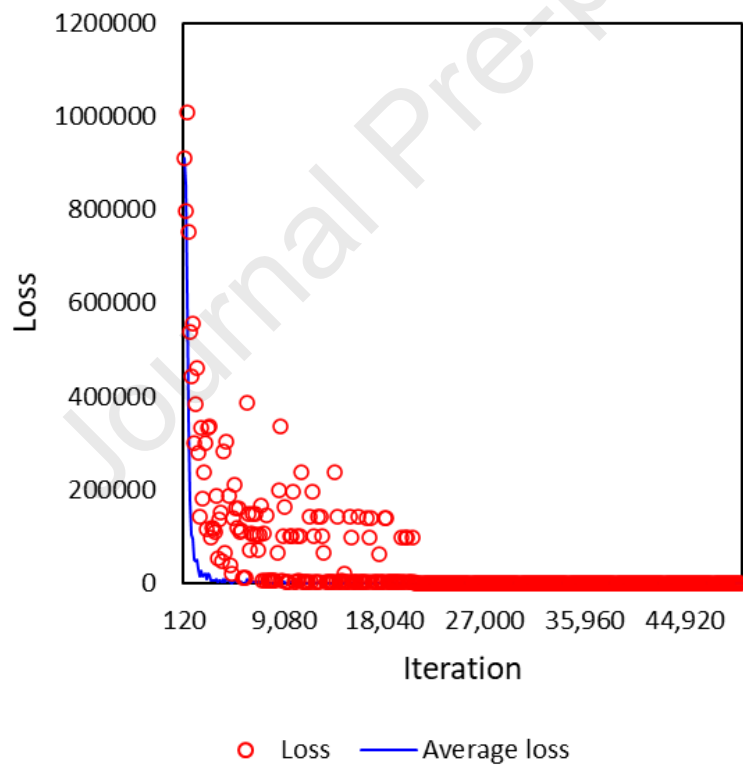
Maintenance activities	Speed (km/h)	Duration for a section (h)	Power (kW)	Energy consumption (kWh)	CO <sub>2</sub> e from Electricity (kg CO <sub>2</sub> e)	Type of material	Material use	CO <sub>2</sub> e from material (kg CO <sub>2</sub> e)	Total CO <sub>2</sub> e (kg CO <sub>2</sub> e)
Tamping	56 [41]	5.36E-06	130.5 [41]	6.99E-04	3.56E-04	-	-	-	3.56E-04
Rail grinding	8 [42]	3.75E-05	120.0 [42]	4.50E-03	2.29E-03	-	-	-	2.29E-03
Ballast cleaning	48 [43]	6.25E-06	838.9 [43]	5.24E-03	2.67E-03	Stone	2,900 kg [43]	6.16E-06	2.67E-03
Sleeper replacement	48 [44]	6.25E-06	36.0 [44]	2.25E-04	1.15E-04	Concrete	1 set	8.68	8.68E+00
Rail replacement	160 [45]	1.88E-06	20.0 [45]	3.75E-05	1.91E-05	Steel	0.6 m	66.6	6.66E+01
Fastening components replacement	-	-	-	-	-	Steel	2 sets (19.48 kg) [46]	36.04	3.60E+01
Ballast unloading	5 [47]	6.00E-05	1491.4 [47]	8.95E-02	4.55E-02	-	-	-	4.55E-02



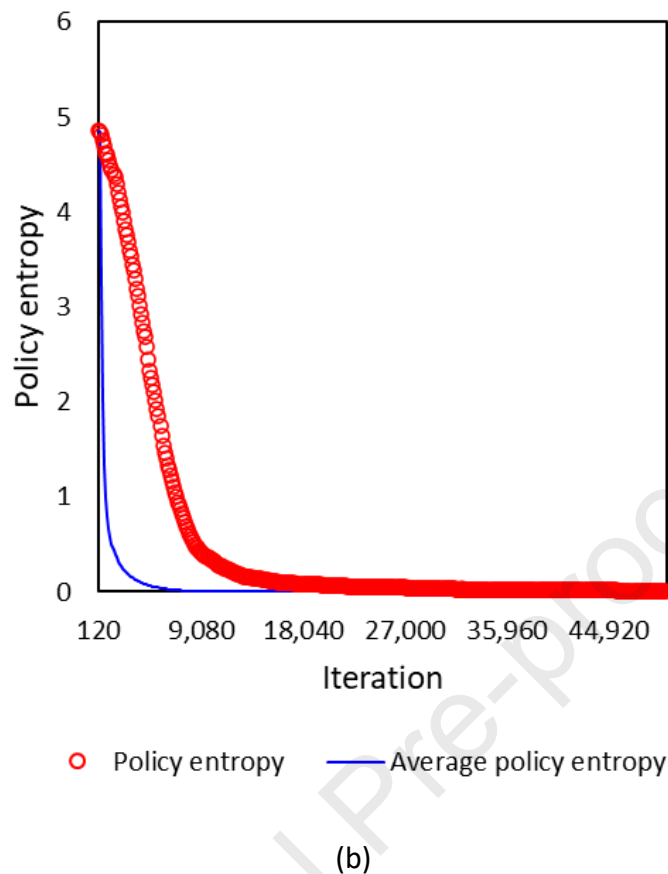
## 412 4 RESULTS AND DISCUSSION

413 4.1 *RL model training*

414 The RL model has been coded to train for 100,000 iterations or training cycles. Python is used  
415 to create the RL model. PPO was utilized to develop the RL model. Figure 7 presents the  
416 training's loss and policy entropy. When the training is completed, the expected loss and  
417 policy entropy should be close to 0. It can be observed that after 20,000 epochs of training,  
418 the loss is near 0, indicating that the RL model has been optimized and the training is  
419 complete. The iteration is shown in the figure through 40,000 to provide a clear view.



(a)

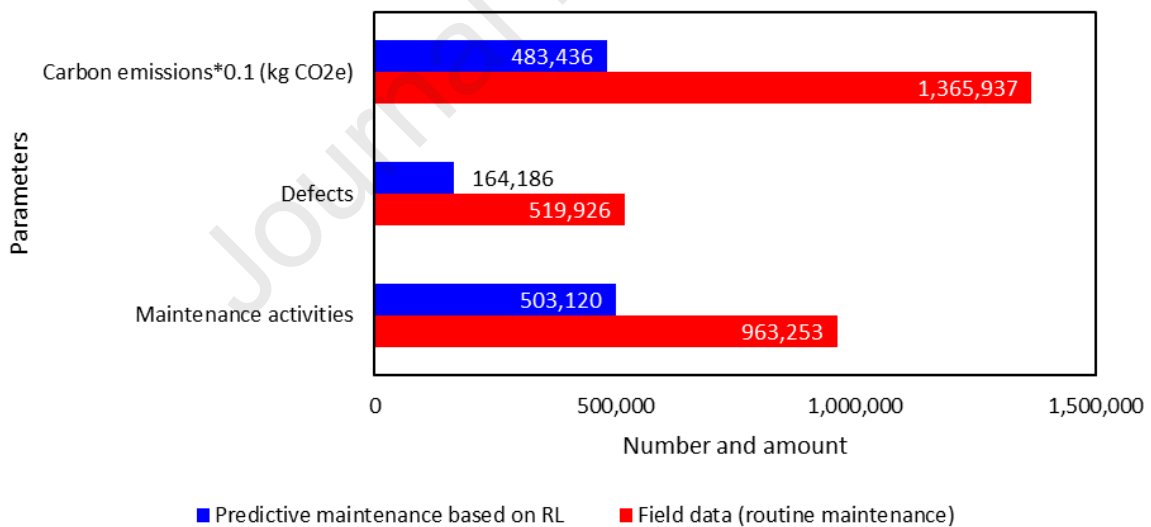


420 Figure 7 RL model training progress (a) loss and (b) policy entropy

#### 421 4.2 RL model performance

422 Figure 8 (a) illustrates the variations in the number of maintenance activities conducted,  
 423 defects that occurred, and carbon emission using field data or routine maintenance and the  
 424 RL model that can be called predictive maintenance. Figure 8 (b)-(d) demonstrates the  
 425 distribution of the number of defects, maintenance activities, and cumulative carbon  
 426 emissions along track sections respectively. The number of maintenance activities performed  
 427 from the field data is 963k, whereas the number of maintenance activities performed through  
 428 RL is 503k. In that instance, using the RL model to make a maintenance decision can reduce  
 429 the number of maintenance tasks by 48%. The number of occurring defects from the field  
 430 data is roughly 520k, whereas the amount from the RL model is decreased to 164k. The  
 431 calculation shows that it reduces the number of defects by 68%. In terms of carbon emission,

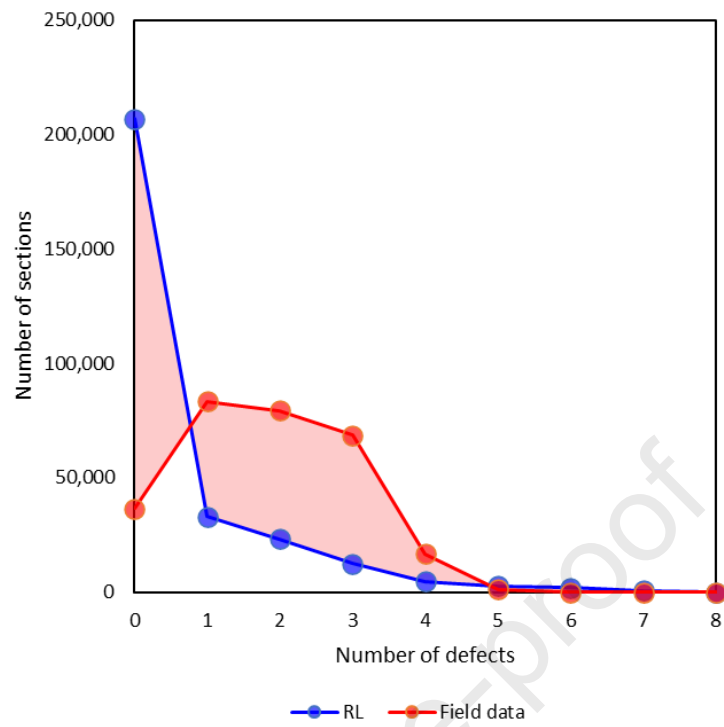
432 performed maintenance activities based on filed data emit 13.7k tons CO<sub>2</sub>e, whereas  
433 performed maintenance activities based on RL emit just 4.83k tons CO<sub>2</sub>e. This suggests that  
434 the new technique can decrease carbon emission from railway maintenance by up to 65%.  
435 According to the findings, the created RL model significantly enhances overall maintenance  
436 efficiency in terms of operation and environment. This is consistent with the preliminary  
437 analysis's finding that maintenance was not carried out efficiently. Some track sections  
438 necessitate more attention while others require less. When maintenance is performed  
439 properly, efficiency may be greatly increased. It can be inferred that the proposed RL model  
440 may greatly reduce carbon emission from railway maintenance operations, which is  
441 consistent with the study's goal.



442

443

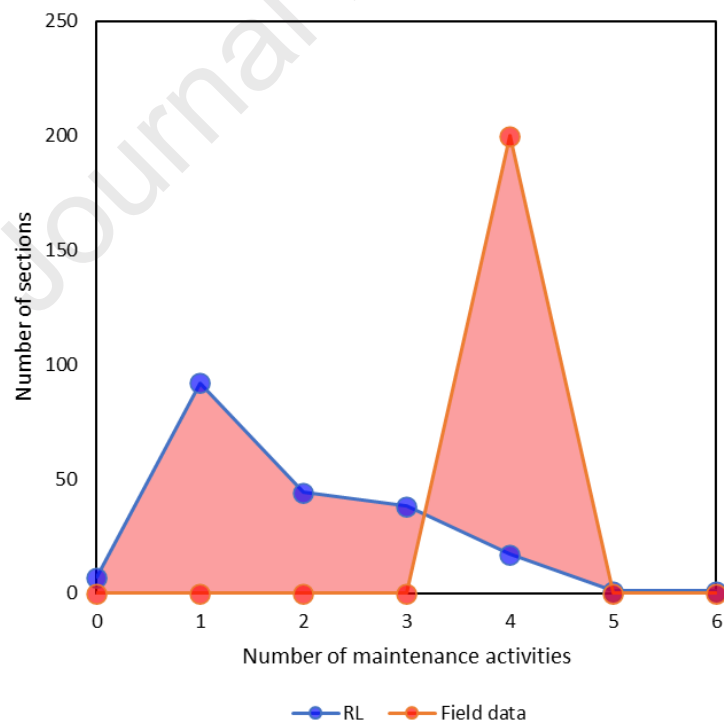
(a)



444

445

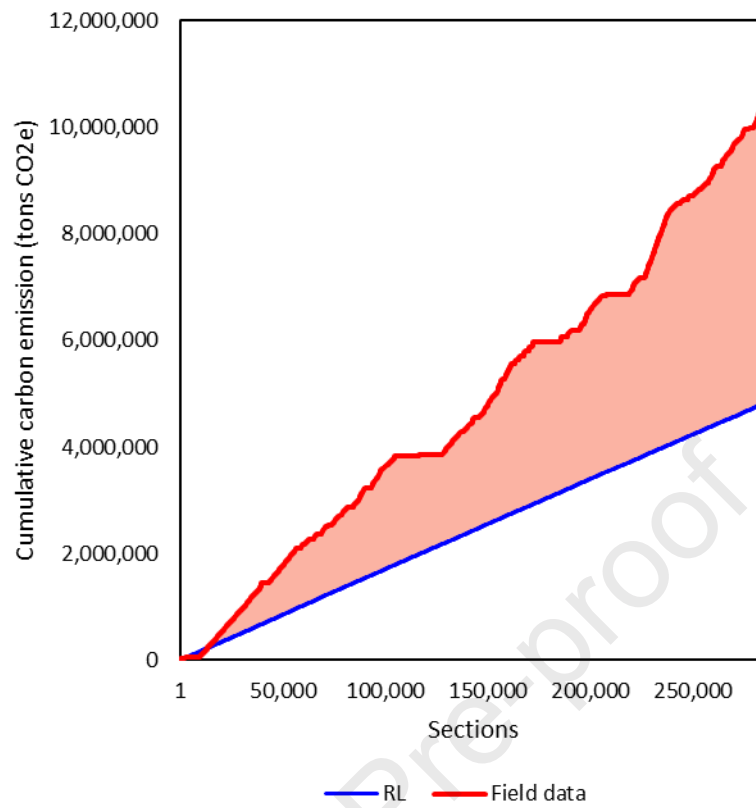
(b)



446

447

(c)



448

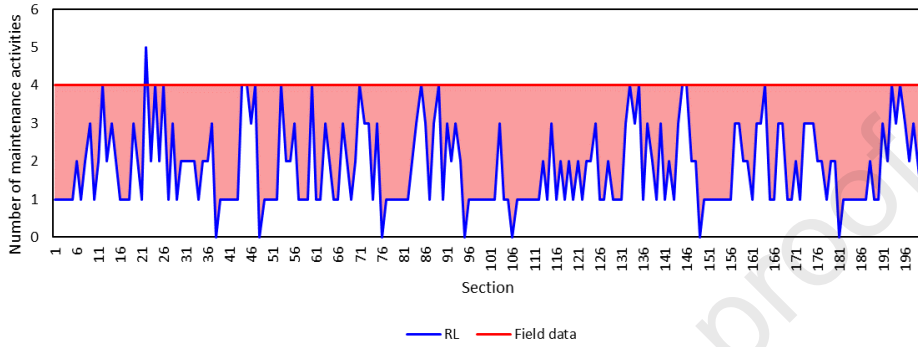
449

(d)

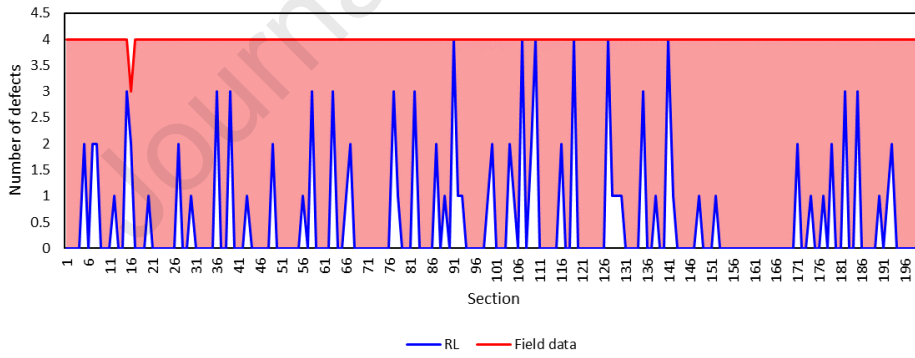
450 Figure 8 Comparison of overall results between field data and RL model (a) overall result, (b)  
451 number of defects distribution, (c) number of maintenance activities distribution, and (d)  
452 cumulative carbon emission along sections

453

454 4.3 Examples of RL model application



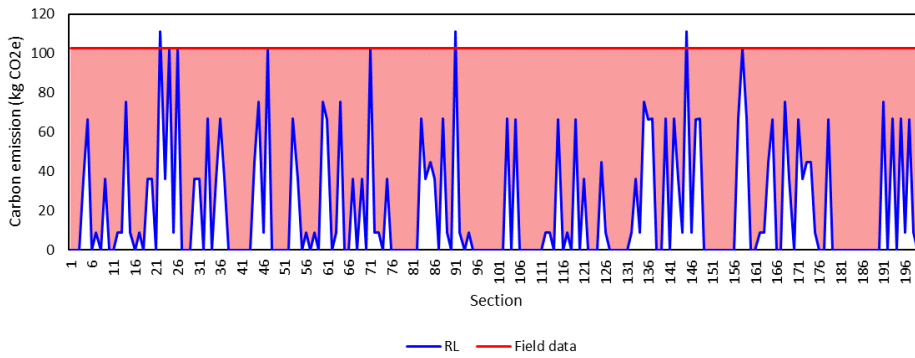
(a)



455

456

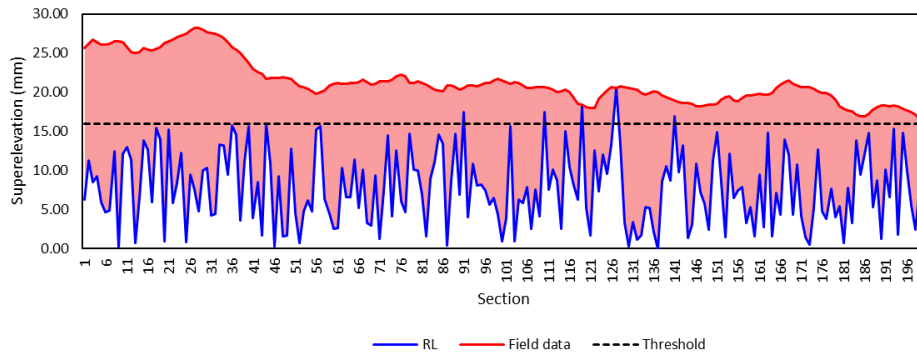
(b)



457

458

(c)



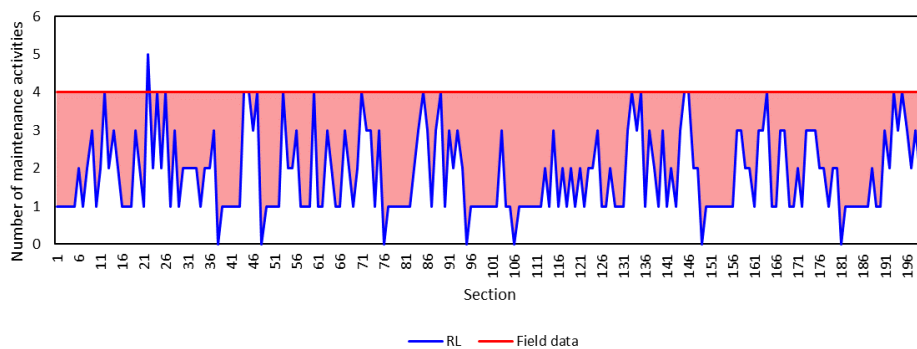
459

460

(d)

461 Figure 9 shows samples of the first 200 track sections. It is clear that the RL model reduces  
 462 the overall amount of maintenance activities, defects, and carbon emission. However, due to  
 463 the stochastic nature of the railway system, the number from the RL model is occasionally  
 464 larger than the field data. Although certain track sections are maintained, defects do develop,  
 465 which poses a challenge in the railway industry. Some part of track geometry parameters is  
 466 demonstrated using super-elevation as an example. According to the maintenance guide, the  
 467 super-elevation threshold is 16 mm. All values from the first 200 sections of the field data  
 468 surpass the threshold, which is unsatisfactory. However, the results of the RL model  
 469 demonstrate that the majority of the super-elevation is within an acceptable range, while  
 470 there are a few sections where the values surpass the threshold, which can occur.

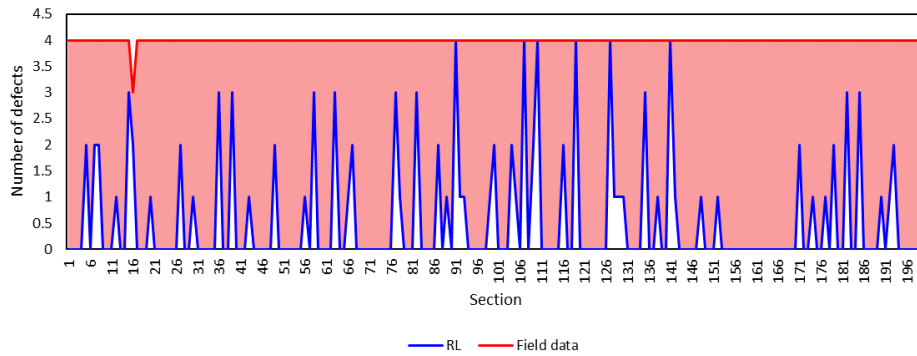
471



472

473

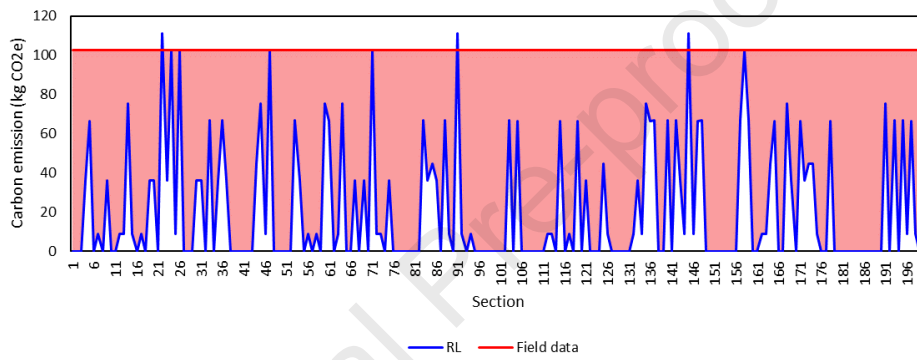
(a)



474

475

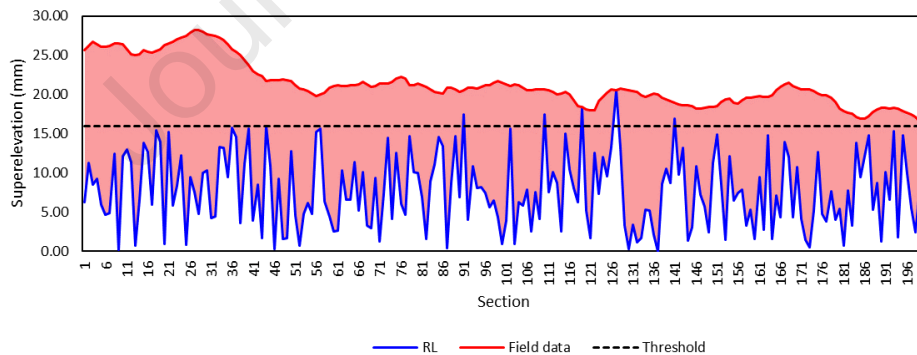
(b)



476

477

(c)



478

479

(d)

480 Figure 9 Examples of differences between field data and RL model (a) maintenance  
 481 activities, (b) defects, (c) carbon emissions, and (d) superelevation

482 From Table 1, rail and fastening component replacements generate the highest carbon  
 483 emission. The use of RL can reduce the number of these two activities by 17% and 72%  
 484 respectively. For more detail, rail replacement is planned to perform while rail grinding can



485 also be performed to improve track quality and causes less carbon emission. At the same  
486 time, fastening component replacement might not be performed efficiently because they are  
487 not defective. These are examples of how RL can improve maintenance efficiency and reduce  
488 carbon emissions based on what RL learns from data. Railway operators can apply the  
489 approach proposed in this study with their data to improve maintenance efficiency by  
490 inputting the current status or parameters of the railway infrastructure and using the  
491 developed RL model to prepare maintenance plans. From the results, the RL model can reduce  
492 carbon emissions, the number of defects, and maintenance costs. It is believed that they will  
493 be able to reduce defects, maintenance costs, and carbon emissions because RL learns from  
494 data and is not biased.

## 495 5 CONCLUSION

496 The objective of this study is to use the RL model to minimize carbon emission from railway  
497 maintenance. The PPO approach was utilized to create the RL model. The states utilized to  
498 train the agent are made up of 12 parameters extracted from track geometry parameters to  
499 track component defects. The action spaces for the RL agent are created by combining seven  
500 maintenance activities. Rewards are dependent on the carbon emission from maintenance  
501 activities undertaken and defects discovered in track sections. These characteristics make the  
502 model developed in this study novel and unique. As a result, the RL agent is trained to  
503 minimize carbon emission while maintaining defect-free railway tracks. Field data obtained  
504 from the 30km track between 2016 and 2019 was utilized to create the RL model.

505 The proposed RL model can achieve the study's goal. It highlights the possibility of lowering  
506 carbon emission from railway maintenance by decreasing maintenance efforts while limiting

507 defects that develop. The findings of the RL model demonstrate that it may cut maintenance  
508 activities by 48%, defects by 68%, and carbon emission by 65% when compared to field data  
509 which is a significant improvement that has never been achieved by using other techniques.  
510 Moreover, contributions of the developed RL model besides the carbon emission and defect  
511 reduction are it can improve the reliability and serviceability of train services because it  
512 reduces the probability of system failure, enhance maintenance and asset management,  
513 reduce environmental impacts, improve resource allocations, improves the safety of  
514 passengers and railway staff, or integrate new technology and support the autonomous  
515 system. In conclusion, the developed RL model can resolve the pain points of railway  
516 maintenance. Railway maintenance can be complex and complicated. Corrective  
517 maintenance is not efficient because the system has to fail first before being fixed. Preventive  
518 maintenance is sometimes too much in terms of maintenance. The developed model can  
519 make predictive maintenance feasible and efficient. Maintenance can be done based on data  
520 not feeling or experience which can be biased. Therefore, it can be tracked and improved over  
521 time. It is worth noting that the maintenance cost has not been considered in this study due  
522 to the confidentiality issue. However, including this part of the data can absolutely improve  
523 the realism of the study.

524 This study will assist railway decision-making groups in better track inspection and  
525 maintenance schedules. Using the methodology provided by this study, they may use their  
526 database data to train the RL model. The produced model may then be used to assist or even  
527 drive selections, which is the ultimate objective of the data-driven concept. Some restrictions,  
528 such as cost, machine, or human resource limitations, can be added or adjusted in the  
529 environment to satisfy their circumstances. Railway operators can apply the developed RL

530 models by inputting the geometry data, maintenance records, and defect inspection. Then,  
531 the models are used to identify maintenance activities that should be performed to maintain  
532 tracks in good condition with the lowest carbon emission. Then, the models will update  
533 themselves and be ready to receive new data from the following years or stages before  
534 identifying the next proper actions continuously. The degree and priority of maintenance can  
535 also be put in action spaces to provide various states that will be interesting for the next stage  
536 of the investigation. Furthermore, the following inspection and measurement plan might be  
537 determined by the current track conditions and maintenance activities. This will undoubtedly  
538 represent real-world applicability, but it will also complicate the study. It is, nonetheless,  
539 intriguing and has the potential to enhance reinforcement learning to mimic real-world  
540 events as closely as feasible.

#### 541 6 AUTHOR CONTRIBUTIONS

542 **Sakdirat Kaewunruen:** Conceptualization, Methodology, Validation, Resources, Data  
543 Curation, Supervision, Funding Acquisition. **Jessada Sresakoolchai:** Conceptualization,  
544 Methodology, Software, Validation, Formal analysis, Investigation, Writing - Original Draft,  
545 Writing - Review & Editing, Visualization.

#### 546 7 DECLARATION OF INTERESTS

547 The authors declare no competing interests.

#### 548 8 ACKNOWLEDGMENT

549 The authors also wish to thank the European Commission for the financial sponsorship of the  
550 H2020-RISE Project no. 691135 "RISEN: Rail Infrastructure Systems Engineering Network",

551 which enables a global research network that addresses the grand challenge of railway  
 552 infrastructure resilience and advanced sensing in extreme environments ([www.risen2rail.eu](http://www.risen2rail.eu)).

## 553 9 REFERENCES

- 554 1. Swe, T.M., P. Jongvivatsakul, and W. Pansuk, *Properties of pervious concrete aiming for*  
 555 *LEED green building rating system credits*. Eng. J., 2016. **20**(2): p. 61-72.
- 556 2. Frelich, L.E. and P.B. Reich, *Will environmental changes reinforce the impact of global*  
 557 *warming on the prairie-forest border of central North America?* Front. Ecol. Environ., 2010.  
 558 **8**(7): p. 371-378.
- 559 3. Houghton, J., *Global warming*. Rep. Prog. Phys., 2005. **68**(6): p. 1343.
- 560 4. Andersson, E., O. Fröidh, S. Stichel, T. Bustad, and H. Tengstrand, *Green Train: concept and*  
 561 *technology overview*. Int. J. Rail Transp., 2014. **2**(1): p. 2-16.
- 562 5. Rungskunroch, P., Z.-J. Shen, and S. Kaewunruen, *Getting it right on the policy prioritization*  
 563 *for rail decarbonization: Evidence from whole-life CO<sub>2e</sub> emissions of railway systems*.  
 564 *Frontiers in Built Environment*, 2021: p. 60.
- 565 6. Szwedo, J.D. *Preventive, predictive and corrective maintenance*.
- 566 7. Mohammadi, R. and Q. He, *A deep reinforcement learning approach for rail renewal and*  
 567 *maintenance planning*, in *Reliability Engineering & System Safety*. 2022, Elsevier. p. 108615.
- 568 8. Yarram, S. and V.V. Krishna *COMPARISON OF REINFORCEMENT LEARNING*  
 569 *ALGORITHMS*. 2022.
- 570 9. Network Rail, *Reliable and resilient track geometry*. 2022.
- 571 10. Network Rail. *Challenge statements*. 2022; Available from:  
 572 [https://www.networkrail.co.uk/industry-and-commercial/research-development-and-](https://www.networkrail.co.uk/industry-and-commercial/research-development-and-technology/research-and-development-programme/challenge-statements/)  
 573 [technology/research-and-development-programme/challenge-statements/](https://www.networkrail.co.uk/industry-and-commercial/research-development-and-technology/research-and-development-programme/challenge-statements/).
- 574 11. Sutton, R.S. and A.G. Barto, *Reinforcement learning: An introduction*. 2018: MIT press.
- 575 12. Zhu, Y., H. Wang, and R.M.P. Goverde. *Reinforcement learning in railway timetable*  
 576 *rescheduling*. 2020. IEEE.
- 577 13. Li, Y., *Reinforcement learning applications*. arXiv preprint arXiv:1908.06973, 2019.
- 578 14. Luong, N.C., D.T. Hoang, S. Gong, D. Niyato, P. Wang, Y.-C. Liang, and D.I. Kim,  
 579 *Applications of deep reinforcement learning in communications and networking: A survey*.  
 580 *IEEE Commun. Surv. Tutor.*, 2019. **21**(4): p. 3133-3174.
- 581 15. Mahmud, M., M.S. Kaiser, A. Hussain, and S. Vassanelli, *Applications of deep learning and*  
 582 *reinforcement learning to biological data*. *IEEE Trans. Neural Netw. Learn. Syst.*, 2018. **29**(6):  
 583 p. 2063-2079.
- 584 16. Zhang, Z., D. Zhang, and R.C. Qiu, *Deep reinforcement learning for power system*  
 585 *applications: An overview*. *CSEE J. Power Energy Syst.*, 2019. **6**(1): p. 213-225.
- 586 17. Kormushev, P., S. Calinon, and D.G. Caldwell, *Reinforcement learning in robotics:*  
 587 *Applications and real-world challenges*. *Robotics*, 2013. **2**(3): p. 122-148.
- 588 18. Abdulhai, B. and L. Kattan, *Reinforcement learning: Introduction to theory and potential for*  
 589 *transport applications*. *Can. J. Civ. Eng.*, 2003. **30**(6): p. 981-991.
- 590 19. Zhou, S.K., H.N. Le, K. Luu, H.V. Nguyen, and N. Ayache, *Deep reinforcement learning in*  
 591 *medical imaging: A literature review*. *Med Image Anal*, 2021. **73**: p. 102193.
- 592 20. Kolm, P.N. and G. Ritter, *Modern perspectives on reinforcement learning in finance*. *Modern*  
 593 *Perspectives on Reinforcement Learning in Finance* (September 6, 2019). *The Journal of*  
 594 *Machine Learning in Finance*, 2020. **1**(1).
- 595 21. Andriotis, C.P. and K.G. Papakonstantinou, *Managing engineering systems with large state*  
 596 *and action spaces through deep reinforcement learning*. *Reliab. Eng. Syst. Saf.*, 2019. **191**: p.  
 597 106483.
- 598 22. Sedghi, M., O. Kauppila, B. Bergquist, E. Vanhatalo, and M. Kulahci, *A taxonomy of railway*  
 599 *track maintenance planning and scheduling: A review and research trends*. *Reliability*  
 600 *Engineering & System Safety*, 2021. **215**: p. 107827.

- 601 23. Šemrov, D., R. Marsetič, M. Žura, L. Todorovski, and A. Srdic, *Reinforcement learning*  
602 *approach for train rescheduling on a single-track railway*. *Transp. Res. B: Methodol.*, 2016.  
603 **86**: p. 250-267.
- 604 24. Cui, Y., U. Martin, and W. Zhao, *Calibration of disturbance parameters in railway operational*  
605 *simulation based on reinforcement learning*. *J. Rail Transp. Plan. Manag.*, 2016. **6**(1): p. 1-12.
- 606 25. Khadilkar, H., *A scalable reinforcement learning algorithm for scheduling railway lines*. *IEEE*  
607 *trans Intell Transp Syst*, 2018. **20**(2): p. 727-736.
- 608 26. Wang, H., Z. Han, Z. Liu, and Y. Wu, *Deep Reinforcement Learning Based Active Pantograph*  
609 *Control Strategy in High-Speed Railway*. *IEEE Trans. Veh. Technol.*, 2022.
- 610 27. Xu, J. and B. Ai, *Experience-driven power allocation using multi-agent deep reinforcement*  
611 *learning for millimeter-wave high-speed railway systems*. *IEEE trans Intell Transp Syst*, 2021.
- 612 28. Zhu, F., Z. Yang, F. Lin, and Y. Xin, *Decentralized cooperative control of multiple energy*  
613 *storage systems in urban railway based on multiagent deep reinforcement learning*. *IEEE*  
614 *Trans. Power Electron.*, 2020. **35**(9): p. 9368-9379.
- 615 29. Deng, K., Y. Liu, D. Hai, H. Peng, L. Löwenstein, S. Pischinger, and K. Hameyer, *Deep*  
616 *reinforcement learning based energy management strategy of fuel cell hybrid railway vehicles*  
617 *considering fuel cell aging*. *Energy Convers. Manag.*, 2022. **251**: p. 115030.
- 618 30. Zhong, J., Z. Liu, H. Wang, W. Liu, C. Yang, Z. Han, and A. Núñez, *A looseness detection*  
619 *method for railway catenary fasteners based on reinforcement learning refined localization*.  
620 *IEEE Trans. Instrum. Meas.*, 2021. **70**: p. 1-13.
- 621 31. Gao, T., Z. Li, Y. Gao, P. Schonfeld, X. Feng, Q. Wang, and Q. He, *A deep reinforcement*  
622 *learning approach to mountain railway alignment optimization*. *Comput.-Aided Civ.*  
623 *Infrastruct. Eng.*, 2022. **37**(1): p. 73-92.
- 624 32. Mohammadi, R. and Q. He, *A deep reinforcement learning approach for rail renewal and*  
625 *maintenance planning*. *Reliability Engineering & System Safety*, 2022. **225**: p. 108615.
- 626 33. Kaelbling, L.P., M.L. Littman, and A.W. Moore, *Reinforcement learning: A survey*. *J Artif*  
627 *Intell Res*, 1996. **4**: p. 237-285.
- 628 34. Alibabaei, K., P.D. Gaspar, E. Assunção, S. Alirezazadeh, T.M. Lima, V.N.G.J. Soares, and  
629 J.M.L.P. Caldeira, *Comparison of on-policy deep reinforcement learning A2C with off-policy*  
630 *DQN in irrigation optimization: A case study at a site in Portugal*. *Computers*, 2022. **11**(7): p.  
631 104.
- 632 35. MRS Logística S.A., *Track geometry thresholds*. 2019.
- 633 36. Kaewunruen, S., J. Sresakoolchai, and Y.-h. Lin, *Digital twins for managing railway*  
634 *maintenance and resilience*. *ERJ Open Res.*, 2021. **1**(91): p. 91.
- 635 37. Kaewunruen, S., M. AbdelHadi, M. Kongpuang, W. Pansuk, and A.M. Remennikov, *Digital*  
636 *Twins for Managing Railway Bridge Maintenance, Resilience, and Climate Change*  
637 *Adaptation*. *Sensors*, 2022. **23**(1): p. 252.
- 638 38. Meddah, M.S. *Recycled aggregates in concrete production: engineering properties and*  
639 *environmental impact*. 2017. EDP Sciences.
- 640 39. Rempelos, G., J. Preston, and S. Blainey, *A carbon footprint analysis of railway sleepers in the*  
641 *United Kingdom*. *Transp Res D Transp Environ*, 2020. **81**: p. 102285.
- 642 40. Hall, J., *Cleaning Up The Steel Industry: Reducing CO2 Emissions with CCUS*. 2021.
- 643 41. Networkrail, *Track treatment fleet*. 2022.
- 644 42. Pouget, *MACHINES FOR TRACK WORK*. 2022.
- 645 43. Progressrail, *KERSHAW® KSC2000 SHOULDER CLEANER*. 2022.
- 646 44. Geismar, *SLEEPER CHANGING MACHINE*. 2022.
- 647 45. Vossloh, *Rail Replacement Wagon SWW*. 2022.
- 648 46. Anyang General International Co., L., *Rail Fastening System*. 2022.
- 649 47. European Patent Office, *Wagon for unloading ballast on a railway*. 2007.

# Interactive reinforcement learning innovation to reduce carbon emissions in railway infrastructure maintenance

Jessada Sresakoolchai

*School of Civil Engineering, University of Birmingham, B15 2TT, Birmingham, United Kingdom*

Sakdirat Kaewunruen\*

*School of Civil Engineering, University of Birmingham, B15 2TT, Birmingham, United Kingdom*

\*Email: [s.kaewunruen@bham.ac.uk](mailto:s.kaewunruen@bham.ac.uk)

\*Tel.: +44 (0) 121 414 2670

## HIGHLIGHTS

- The first reinforcement learning model for railway carbon emissions reduction
- Field data robustly enables the creation of customized environments for the model
- Environment's states are obtained from defective track geometry and track component
- A complex combination of maintenance activities is adopted as an action space
- Reduce the defect and carbon emissions using an interactive and dynamic approach

## Key novelties:

This is an original paper, which has neither previously nor simultaneously in whole or in part been submitted anywhere else. The paper presents a world-first novel AI-based decision making tool to help reduce carbon emission in railway maintenance, which is one of the primary contributors to global warming throughout the whole lifecycle of the critical asset and infrastructures. The new data-driven method capable of co-simulations with other GUI frameworks using real-world datasets is highly innovative, and novel and we have obtained the field data from joint research collaboration with rail industry globally. Not only can this method deal with rail geometry, but it can also effectively manage complex infrastructure component replacement and repairs, driven by sustainability goals. This can be successful with the computer-aided innovation that enables co-simulation. This outcome will lead to the better understanding and implementation of novel data-driven decision making framework for environmental-friendly maintenance of infrastructures, which are very critical to sustainable development worldwide.

**Declaration of interests**

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Journal Pre-proof